CLAIMS

1. A method for morphological disambiguation, comprising:

receiving an input string;

morphologically analyzing the string to generate a list of candidate analyses of the string, each candidate analysis comprising a respective word and a linguistic pattern of the word; and

evaluating the pattern of each of the analyses against a predefined criterion in order to select one or more of the analyses from the list.

2. A method according to claim 1, wherein receiving the input string comprises receiving a word in a Semitic language.

3. A method according to claim 2, wherein the Semitic language comprises Hebrew.

4. A method according to claim 1, wherein the linguistic pattern comprises a specification of at least one characteristic of the word, selected from a set of characteristics including a part of speech, prefix, number, gender and person of the word.

5. A method according to claim 4, wherein the specification of the at least one characteristic comprises a specification of all of the characteristics in the set.

6. A method according to claim 5, wherein when the base word comprises a verb, the linguistic pattern further comprises a designation of a tense and conjugation pattern of the verb.

7. A method according to claim 1, wherein each of the analyses has a lemma and a paradigm determined by the word and the linguistic pattern thereof, and wherein evaluating the pattern comprises eliminating one of the analyses from the list if it has the same lemma and paradigm as another of the analyses.

8. A method according to claim 1, wherein evaluating the pattern comprises determining a relative frequency of occurrence of the pattern of each of the analyses, and selecting the at least one of the analyses whose frequency of occurrence is above a predetermined threshold.

9. A method according to claim 8, wherein determining the relative frequency of occurrence comprises morphologically analyzing a corpus of text and finding the frequency of occurrence of the pattern in the corpus.

10. A method according to claim 9, wherein determining the relative frequency of occurrence comprises storing in a table the frequency of occurrence found in the corpus, and looking up the pattern in the table.

11. A method according to claim 8, wherein selecting the at least one of the analyses comprises setting the threshold so as to control how many of the analyses from the list are selected.

12. A method according to claim 8, wherein selecting the at least one of the analyses comprises selecting the at least one of the analyses based on the pattern thereof, and substantially independently of the respective word.

13. A method according to claim 1, and comprising searching in a corpus of text for a match to the input string using the one or more selected analyses.

14. A method according to claim 1, and comprising checking for spelling errors in the input string using the one or more selected analyses.

15. A method for searching a corpus of text made up of words, comprising:

morphologically analyzing the words in the corpus to generate, for each of at least some of the words, a list of candidate analyses, each candidate analysis comprising a respective lemma and a linguistic pattern relating the lemma to the analyzed word;

evaluating the pattern of each of the analyses against a predefined criterion in order to select one or more of the analyses from the list for each of the analyzed words;

entering the lemmas of the selected analyses in an index of the corpus; and

applying a search query to the index.

16. A method according to claim 15, wherein applying the search query comprises:

receiving an input text string;

morphologically analyzing and disambiguating the string to generate one or more search lemmas for the string; and

comparing the search lemmas to the index.

17. A method according to claim 15, wherein the words in the corpus comprise words in a Semitic language.

18. A method according to claim 17, wherein the Semitic language comprises Hebrew.

19. A method according to claim 15, wherein the linguistic pattern comprises a specification of at least one characteristic of the word, selected from a set of characteristics including a part of speech, prefix, number, gender and person of the word.

20. A method according to claim 15, wherein evaluating the pattern comprises determining a relative frequency of occurrence of the pattern of each of the analyses, and selecting the at least one of the analyses whose frequency of occurrence is above a predetermined threshold.

21. A method according to claim 20, wherein selecting the at least one of the analyses comprises selecting the at least one of the analyses based on the pattern thereof, and substantially independently of the respective word.

22. A computer software product, comprising a computer-readable medium in which program instructions are stored, which instructions, when read by a computer, cause the computer to morphologically analyze an input string to generate a list of candidate analyses of the string, each candidate analysis comprising a respective word and a linguistic pattern of the word, and to evaluate the pattern of each of the analyses against a predefined criterion in order to select one or more of the analyses from the list.

23. A product according to claim 22, wherein the input string comprises a word in a Semitic language.

24. A product according to claim 23, wherein the Semitic language comprises Hebrew.

25. A product according to claim 22, wherein the linguistic pattern comprises a specification of at least one characteristic of the word, selected from a set of characteristics including a part of speech, prefix, number, gender and person of the word.

26. A product according to claim 22, wherein the instructions further cause the computer to search in a corpus of text for a match to the input string using the one or more selected analyses.

27. A computer software product, comprising a computer-readable medium in which program instructions are stored, which instructions, when read by a computer, cause the computer to morphologically analyze the words in the corpus to generate, for each of at least some of the words, a list of candidate analyses, each candidate analysis comprising a respective lemma and a linguistic pattern relating the lemma to the analyzed word, to evaluate the pattern of each of the analyses against a predefined criterion in order to select one or more of the analyses from the list for each of the analyzed words, to enter the lemmas of the selected analyses in an index of the corpus, and to apply a search query to the index.

28. A product according to claim 27, wherein the instructions further cause the computer to receive an input text string, to morphologically analyze and disambiguate the string to generate one or more search lemmas for the string, and to compare the search lemmas to the index.

29. Apparatus for morphological disambiguation, comprising a linguistic processor, which is adapted to

receive an input string, to morphologically analyze the string to generate a list of candidate analyses of the string, each candidate analysis comprising a respective word and a linguistic pattern of the word, and to evaluate the pattern of each of the analyses against a predefined criterion in order to select one or more of the analyses from the list.

30. Apparatus according to claim 29, wherein the input string comprises a word in a Semitic language.

31. Apparatus according to claim 30, wherein the Semitic language comprises Hebrew.

32. Apparatus according to claim 29, wherein the linguistic pattern comprises a specification of at least one characteristic of the word, selected from a set of characteristics including a part of speech, prefix, number, gender and person of the word.

33. Apparatus according to claim 29, wherein the processor is further adapted to search in a corpus of text for a match to the input string using the one or more selected analyses.

34. Apparatus for searching a corpus of text made up of words, comprising a linguistic processor, which is adapted to morphologically analyze the words in the corpus to generate, for each of at least some of the words, a list of candidate analyses, each candidate analysis comprising a respective lemma and a linguistic pattern relating the lemma to the analyzed word, to evaluate the pattern of each of the analyses against a predefined criterion in order to select one or more of the analyses from the list for each of the analyzed words, to enter the lemmas of the selected analyses in an

index of the corpus, and to apply a search query to the index.

35. Apparatus according to claim 34, wherein the processor is further adapted to receive an input text string, to morphologically analyze and disambiguate the string to generate one or more search lemmas for the string, and to compare the search lemmas to the index.